**Task 1:**

**TITANIC SURVIVAL PREDICTION**

**Introduction:**

**Problem Definition and Objective:**

Use the Titanic dataset to build a model that predicts whether a passenger on the Titanic survived or not. This is a classic beginner project with readily available data. The dataset typically used for this project contains information about individual passengers, such as their age, gender, ticket class, fare, cabin, and whether or not they survived.

**Dataset Description:**

**Overview of the Titanic dataset and key features:**

The Titanic Data set is in the csv format. It consists of 12 columns.

The size of the data set is 891 rows and 12 columns. It also consists of many missing values.

The dataset typically includes the following key columns:

1. Passenger Id: Unique identifier for each passenger.
2. Survived: Target variable (0 = No, 1 = Yes).
3. P class: Ticket class (1st, 2nd, or 3rd class).
4. Name: Passenger name (often not directly used for predictions but can derive features).
5. Sex: Gender (Male or Female).
6. Age: Age of the passenger.
7. Sib/ Sp : Number of siblings/spouses aboard.
8. Par/ch: Number of parents/children aboard.
9. Ticket: Ticket number (can sometimes be used for feature extraction).
10. Fare: Amount paid for the ticket.
11. Cabin: Cabin number (often has missing data).
12. Embarked: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).

**Data Cleaning and Preprocessing:**

1.Handling the missing Values

2. Convert categorical variables to numeric using label encoding or one hot encoding

3.Drop unnecessary columns

**Exploratory Data Analytics:**

1.summary statistics

2. visualizing Survival distribution

3.Analyzing categorical variables

* Survival by Gender
* Survival by passenger Class
* Survival by Embarked

4.Analyzing Numeric variables

5. Correlation matrix

6. Relationship Between multiple variables

* + Survival by Gender and Class
  + Survival by Age and Class

**Feature Engineering:**

1. Create Family Size.
2. Create Is alone
3. Bin age into groups
4. Bins fare into groups

**Model Building:**

1. Split the data into training and testing sets.
2. Train the Machine Learning Models to predict Survival.

**Evaluate the model:**

Assess the model performance using the metrices like accuracy, precision, recall and FI score.

**Fine Tune-the Model:**

Use the hyperparameter Tuning to improve the performance.

Grid search CV for hyperparameter optimization.

**Evaluate the model on the test data:**

Reassess the final model performance using the test data to optimize generalization.

**Analyse the feature importance:**

For models like random Forest, Gradient Boosting analyse which feature are most important for prediction.

**Finally, Save the model with .pkl extension.**